Flood Discharge Prediction By Multiple Linear Regression Method

Case Study: Ciliwung Watershed – Cisadane

Tsamrotul Jannah
Civil Engineering
Gunadarma University
Depok, Indonesia
tsamrotuljannah03@gmail.com

Diyanti
Civil Engineering
Gunadarma University
Depok, Indonesia
diyanti@staff.gunadarma.ac.id

Budi Santosa
Civil Engineering
Gunadarma University
Depok, Indonesia
bsantosa@staff.gunadarma.ac.id

***Abstract*— Flooding is a natural phenomenon that is often encountered in Indonesia, namely the widespread flow of river water, or a lake that soak the mainland so that a method is required to produce predictions accurately. This research aims to predict the flood discharge and find out the correlation between rainfall and discharge. The analytical methods used in this study are multiple linear regression methods with maximum rainfall data on Ranca Bungur, Pasir Jaya, Kracak and Cigudeg rain stations as a dependent variable and maximum discharge data on Ciliwung – Cisadane Watershed as an independent variable. Multiple linear regression equations are formed using rain and discharge data in the years 2004 – 2010, annual rainfall and discharge data for the year 2011 – 2017 are used as a verification of predicted results. The correlation between rainfall and debits is strongly rated due to the average in one year the coefficient of its correlates is 0.864 whereas the average Mean Absolute Percentage Error (MAPE) in one year amounted to 46.617%.**

Keywords—Forecast, Flood, Rainfall, Multiple Linear Regression

I. INTRODUCTION

Correlation and regression both have a very close relationship. Every regression must have a correlation, but the correlation does not necessarily follow regression. Correlation that is not followed by regression is a correlation between two variables that do not have a casual/causal relationship, or relationship functional. To determine whether the two variables have a causal relationship no, it must be based on theory or concepts about two variables.

The rainfall-runoff relationship is an important problem in hydrology and is the most basic component in the water resources evaluation process (Junsawang et al., 2007). The relationship between rainfall and runoff, especially the conversion of rainfall into runoff, in

a river basin is a very complex hydrological phenomenon. Where, this process is non-linear with changing time and spatially distributed (Rajurkar, et.al., 2003).

Rain that falls continuously so that rivers and lakes can no longer hold water can cause flooding. Floods can be caused by very heavy rain with a very large discharge so that eventually flooding occurs because the abundant rainwater cannot immediately flow through the channels or ditches.

The aim of this research is to predict the average flood discharge that will occur in the future and determine the correlation between rainfall at the Ranca Bungur rain station, Pasir Jaya station, Kracak station and Cigudeg station and the flood discharge of the Ciliwung - Cisadane watershed.

II. METHODOLOGY

The analytical method used in this research is multiple linear regression analysis with previously carried out normal testing (Smirnov-Kolmogorov), goodness of fit test, simultaneous test/F test.

The stages carried out in this research are as follows:

1. Data collection
   The data used is maximum rainfall data and maximum discharge data in the Ciliwung-Cisadane watershed.
2. Normality test
   The normality test is carried out to determine whether the data used in the research is normally distributed or not. If the data is not normally distributed then data must be replaced so that the research can continue.
3. Data analysis

Data analysis is a process of processing data obtained through observation. The analysis carried out in this research was multiple linear regression analysis. Regression is a measuring tool to determine whether or not there is a correlation between variables.

4. Validation with field data
Prediction data is matched with field data using Mean Absolute Percentage Error (MAPE).

A flow diagram of the research design is presented, which can be seen in Figure 1.

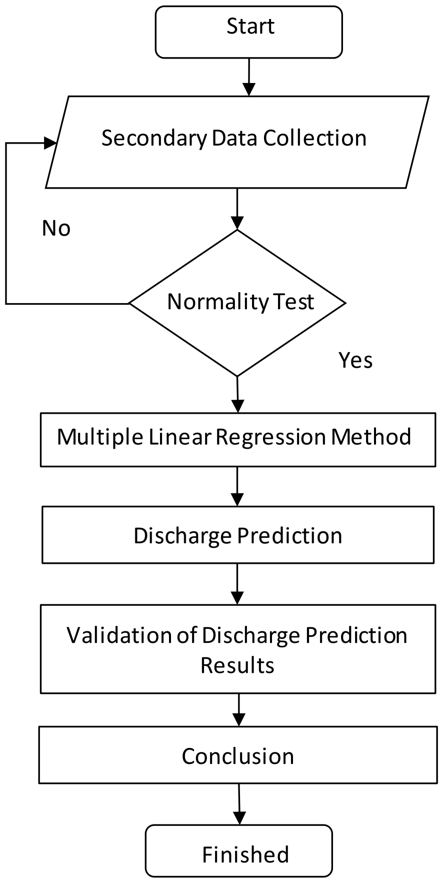

Figure 1 Research Design Flow Diagram

A. Data Collection

The data used in this research is secondary data from measurements obtained from BBWS Ciliwung Cisadane. Data is collected through copies or derivatives of data/copies from relevant agencies through procurement and purchase of data/maps. Apart from that, data is also obtained from internet access.

The type of data used to carry out this research analysis consists of rainfall data in the Cisadane watershed and discharge data.

B. Analysis And Discussion

The analysis and discussion stage contains analysis and discussion of the relationship between rainfall and discharge using the linear regression method. So that it can be known/predicted the flood discharge that will occur in the future.

C. Conclusion

The conclusion stage contains several results obtained based on research, conclusions are made based on research that has been carried out and suggestions that are expected to be used for further research.

III. DATA ANALYSIS

The data used is rainfall data and discharge data in the Ciliwung – Cisadane River Basin (DAS). The Ciliwung – Cisadane watershed is administratively located in West Java Province. Geographically, the Ciliwung – Cisadane watershed is located at 106º 20' 50" – 106º 28' 20" East Longitude and 6º 0' 59" – 6º 47' 02" South Latitude (Ciliwung-Cisadane Watershed Management Center). The rainfall data used is maximum rainfall data from 4 (four) rain stations, namely Ranca Bungur, Pasir Jaya, Kracak and Cigudeg rain stations. The data period used is 2004 – 2017.

Rainfall data is used as a predictor of discharge data, and is divided into 2 (two) periods. The first period 2004 – 2010 was used to build the prediction equation model and the second period 2011 – 2017 to verify the prediction results.

1. Look for discharge data for the Cisadane sub-watershed and other supporting references for the purposes of the analysis to be carried out.
2. Grouping variable x with variable y, both to build a regression equation model and predict river discharge.
3. Test the normality of the data by entering annual rainfall data with monthly discharge in the same year into the SPSS program with the Smirnov-Kolmogorov test to find out whether the data has a normal distribution.
4. Calculate the average variance value and the variance value for each average variance (dXi).
5. Calculate the coefficients for each variable with a system of linear equations using a matrix using the Microsoft Excel program.
6. Calculate the correlation coefficient for the multiple linear regression method on variable x (rainfall at each rain post) as a whole with variable y (observed river discharge).
7. Make debit predictions from the results of the equation model that has been built from the discharge data.

8. Match the debit prediction results using the evaluation number criteria, namely the average absolute percentage error (MAPE).

A. *Multiple Linier Regression*

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

Multiple linear regression is a regression model that explains the relationship between response variables (dependent variables) and factors that influence more than one predictor (independent variables). The multiple linear regression prediction method is carried out by forming a regression equation using more than one independent variable.

For analysis using the regression method, two types of variables are distinguished, namely the dependent variable or response variable and the independent variable. Independent variables (independent) are variables that can influence the dependent variable or variables that can predict the price of the dependent variable.

Multiple linear regression is almost the same as simple linear regression, only in multiple linear regression there is more than one estimator variable. The purpose of multiple linear regression analysis is to measure the intensity of the relationship between two or more variables and make an approximate prediction of the value of Y over X. In general, the multiple linear regression model is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_n X_n + \varepsilon \quad (1)$$

Which $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ are the coefficients or parameters of the model.

IV. NORMALITY TEST

The normality test is carried out to determine whether the distribution value of the data used comes from a normally distributed population or whether the variables have a normal distribution or not. The data normality test was carried out as the first step in carrying out linear regression analysis. The results of the Smirnov-Kolmogorov test analysis can be seen in Table 1.

The normality test is carried out to determine whether the distribution value of the data used comes from a normally distributed population or whether the variables have a normal distribution or not. The data normality test was carried out as the first step in carrying out linear regression analysis. The results of the Smirnov-Kolmogorov test analysis can be seen in Table 1.

TABLE I. SMIRNOV-KOLMOGOROV TEST SPSS PROGRAM

One sample kolmogorov smirnov test		**Unstandardized residuals**
N		12
Normal parametersab	Mean	0E-7
	Std. Deviation	20.82116617
	Absolute	0.107
Most extreme differences	Positive	0.084
	Negative	-0.107
Kolmogorov smirnov z		0.107
Asymp. Sig. (2-tailed)		0.200

Note. a.Test distribution is normal
b. Calculated from data

Table 1 explains that based on the normality test with the Kolmogorov Smirnov Test, a significance value of 0.631 was obtained, thus 0.200>0.005 or p>0.005 so it can be concluded that the data is normally distributed. This means that the assumption of data normality is met.

A. *Regression Method*

The regression method used is the multiple linear regression method. Coefficient calculations were carried out using linear equations with Microsoft Excel. The linear regression equation for January, February, March, April May, June, July, August, September, November and December is as follows:

Y Jan = 1585.1461– 1.97216X1– 7.91170X2 5.53392X3– 5.29128X4 (2)

Y Feb = -128.0532+2.84036X1+0.51652X2+ 4.08069X3+1.1765X4 (3)

Y Mar = 3.09096+6.59689X1– 0.37266X2+1.89902X3+0.15438X4 (4)

Y Apr = 86.91623+075661X1+ 0.78289X2+0.54115X3+ 0.09172X4 (5)

Y May = 76.29722 +1.80240X1+1.39160X2– 0.16684X3+0.09904X4 (6)

Y Jun = 142.86794 +0.85829X1+0.66529X2- 0.65624X3– 0.80054X4 (7)

Y Jul = 68.54616 +1.75382X1+0.58714X2– 0.23971X3+0.75602X4 (8)

Y Aug = 67.99214– 2.69232X1– 0.13275X2+1.58144X3+0.9658X4 (9)

Y Sept = 208.29604– 3.51713X1 – 0.15976X2 +0.5969X3– 0.28191X4 (10)

Y Oct = 200.69488– 5.05248X1 –0.52843X2+ 1.13013X3+ 0.19358X4 (11)

Y Nov = 116.55925 +2.73043X1 + 1.11479X2– 0.23611X3– 0.68757X4 (12)

Y Dec = 120.60265 –0.6365X1 –1.33467X2+ 2.25091X3+0.83764X4 (13)

B. Regression Variety Test/Regression F Test

The F test was carried out to test the simultaneous influence of 4 (four) variables X1, X2, X3, and X4 on discharge (Y) in predicting flood discharge. The results of the F test can be seen in the following table.

TABLE II. FOUR VARIABLE REGRESSION TYPES

Source of Diversity (SK)	Degrees of Free (DB)	Sum of Squares (JK)	Middle Square (KT)	F Count	F Table (5%)
Regression	4	5807.98	1451.99	29.88	19.24
Residual	2	97,170	48.58		
Total	6	5905.15			

Based on the results of the variance analysis above, it turns out that Fcount> F(table 5%) this means that H0 is rejected which states that the simple linear estimator (y) regression line obtained is the best regression line to explain that the independent variable x has a real effect on the variable not free y.

C. Correlation Coeficient and Determination Coefficient

The correlation coefficient and determination coefficient values for each month can be seen in Table 3.

TABLE III. CORRELATION COEFFICIENT AND DETERMINATION COEFFICIENT VALUES

	R	R2
Jan	0.871	0.758
Feb	0.943	0.889
Mar	0.587	0.345
Apr	0.875	0.766
May	0.880	0.744
Jun	0.992	0.984
Jul	0.883	0.779
Aug	0.993	0.871
Sept	0.902	0.813
Oct	0.972	0.944
Nov	0.581	0.337
Dec	0.945	0.893

Based on Table 3, the strongest correlation coefficient (R) between rainfall at the Ranca Bungur, Pasir Jaya, Kracak, and Cigudeg rain stations on discharge in the Ciliwung – Cisadane watershed is in August at 0.993 and the weakest in November at 0.581 with an average coefficient -the average in one year is 0.864.

D. Average Flood Discharge Prediction

Discharge prediction is carried out in addition to predicting the amount of discharge in future years as well as validating/matching the results of the multiple linear regression equation that has been produced in the calculation with the discharge in the field. The flood discharge prediction results are as shown in Figures 2 to 8

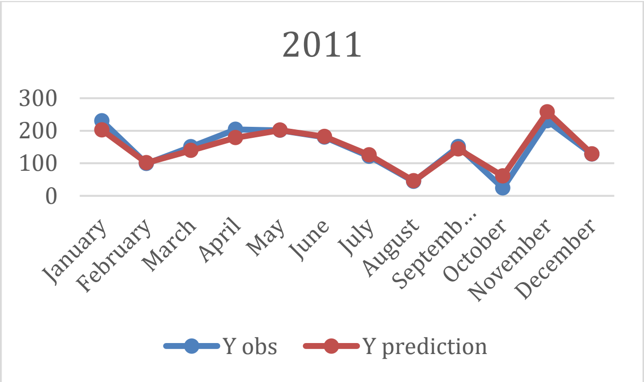

Figure 22011 Average Flood Discharge Prediction Results

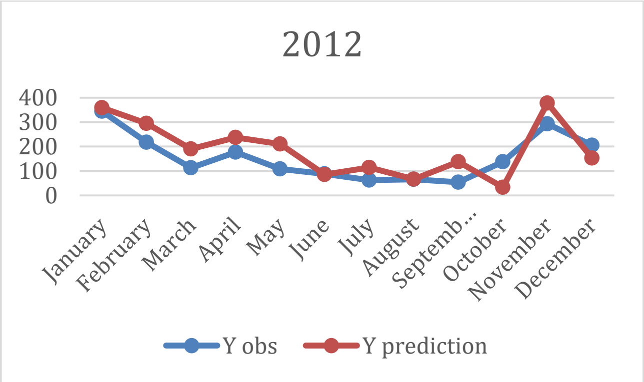

Figure 3 2012 Average Flood Discharge Prediction Results

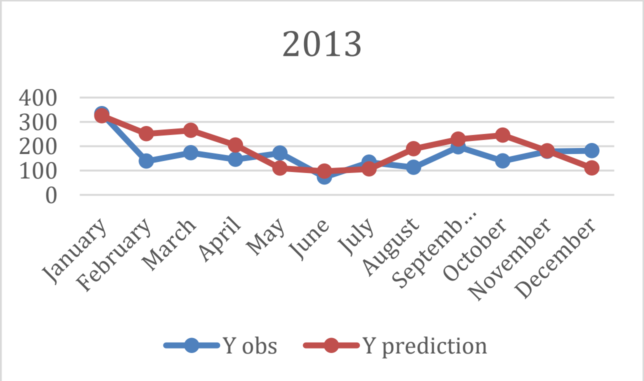

Figure 42013 Average Flood Discharge Prediction Results

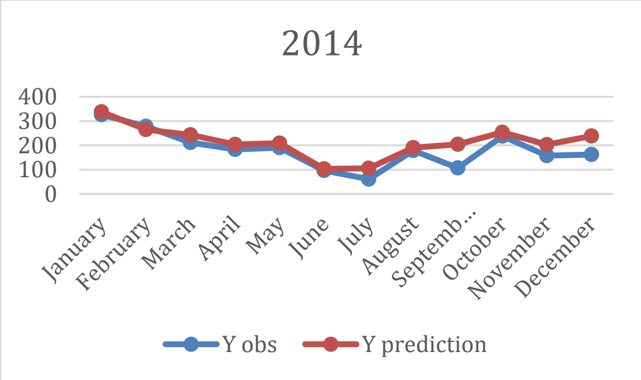

Figure 52014 Average Flood Discharge Prediction Results

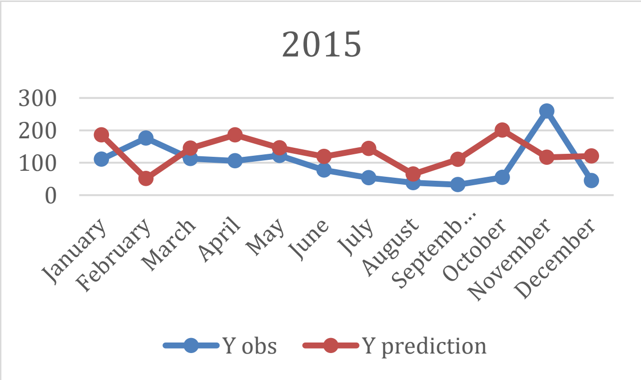

Figure 6 Graph of Average Flood Discharge Prediction for 2015

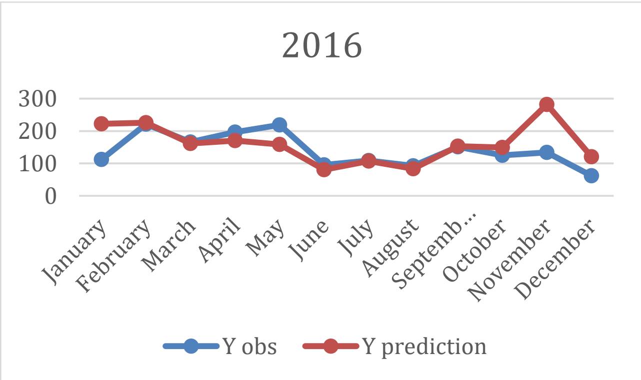

Figure 7 Graph of Average Flood Discharge Prediction for 2016

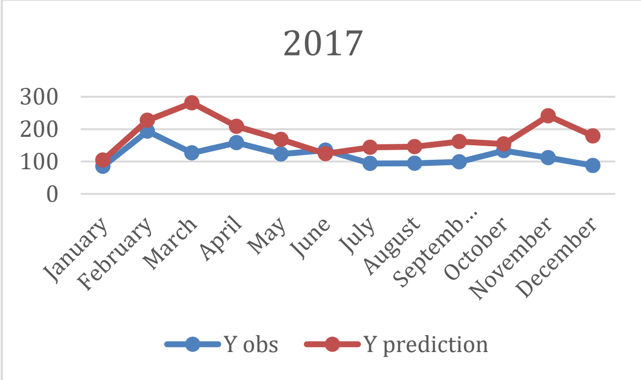

Figure 8 Average Flood Discharge Prediction Graph 2017

V. VERIFY FORECAST RESULT

The average flood discharge prediction results were then verified using the Mean Absolute Percentage Error (MAPE) method to see the level of accuracy of the prediction. The verification results are as follows.

TABLE IV. VERIFICATION RESULTS FROM JANUARY TO DECEMBER

Month	MAPE
Jan	29,989
Feb	30,633
Mar	42,534
Apr	31,198
May	31,802
Jun	17,283
Jul	57,571
Aug	30,228
Sept	81,691
Oct	87,911
Nov	50,115
Dec	68,451

Table 4 shows that the MAPE results for each month are different. The smallest result was in June, namely 17.283%, the largest value was in October, amounting to 87.911%, and the average MAPE value for 1 (one) year was 46.617%.

VI. CONCLUSION

The prediction results of the average flood discharge in 2011, 2012, and 2014 produced quite good predictions because they had a pattern similar to the observed discharge pattern, while the prediction of the average flood discharge in 2013, 2015, 2016, and 2017 produced predictions that were not good because it produces patterns that tend to be different from the observed discharge.

Verification of the prediction results shows that the correlation value is quite significant, the correlation value between rainfall at the four observation stations and the discharge of the Ciliwung - Cisadane watershed has an average correlation coefficient in 1 (one) year of 0.864. The strongest correlation coefficient value is in August, namely 0.993 and the weakest correlation coefficient value is in November, namely 0.581. The percentage of average error value (MAPE) in 1 (one) year is 46.617%.

REFERENCES

[1] Pengelolaan Daerah Aliran Sungai. Gadjah Mada University Press. Yogyakarta.

[2] A.S.K. Kartika, Analisis Prediksi Debit Bulanan Menggunakan Metode Regresi Linier Berganda (Studi Kasus : Sub Das Cisanggarung). Fakultas Teknik Sipil dan Perencanaan Universitas Gunadarma, 2016.

[3] Balai Besar Wilayah Sungai Ciliwung – Cisadane, Laporan Akhir Review Studi Water Balance Sungai Cisadane Tahun Anggaran 2012. Directorate general of water resources Ministry of Public Works and Housing. 2012.

[4] E. Damanhuri. Statistika. Institut Teknologi Bandung. Bandung. 1995.

[5] I. Fajri. Membuat Pola Sebaran Hujan dan Peta Ishoyet pada DAS Ciliwung – Cisadane. Program Studi Teknik Sipil. Fakultas Teknik. Universitas Indonesia. Depok. 2009.

[6] N. Hadisusanto, Aplikasi Hidrologi, . Jogja Mediautama, Yogyakarta. 2010.

[7] W. Handayani, and Y. Indrajaya. "Analisis Hubungan Curah Hujan dan Debit Sub DAS Ngatabaru, Sulawesi Tengah (The Analysis of Rainfall and Discharge Relationship on Ngatabaru Sub Sub Watershed, Central Sulawesi). Balai Penelitian Kehutanan Ciamis. Ciamis. 2011.

[8] N.Q. Hung, M.S. Babel, S. Weesakul, N.K. Tripathi. An Artificial Neural Network Model for Rainfall Forecasting in Bangkok, Thailand. Hydrology and Earth System Science vol. 13 pp. 1413-1425, 2009.

[9] L.I. Grenti. "Peringatan Dini Banjir pada DAS Ciliwung dengan Menggunakan Data Curah Hujan." Fakultas Pertanian. Institut Pertanian Bogor. 2006.

[10] H.N. Indrabayu, M.S. Pallu. Achmad A, Fikha CL. "Prediksi Curah Hujan dengan Jaringan Saraf Tiruan" Prosiding Hasil Penelitian Fakultas Teknik Grup Teknik Elektro Indonesia, December 2012.

[11] E. Junaidi. "Peranan Penerapan Agroforestry Terhadap Hasil Air Daerah Aliran Sungai (DAS) Cisadane (The Role of Agroforestry Implementation to Water Yield in Cisadane Watershed)." Jurnal Penelitian Agroforestry, Vol. 1 No.1 pp. 41 – 53, August 2013.

[12] J.S. Milton and J.C. Arnold. Introduction to Probability and Statistics Principles and Applications for Engineering and The Computing Sciences. McGraw-Hill Publication Co. New York. 1990.

[13] S.W. Mudjonarko. Aplikasi Metode Nakayasu Guna Prediksi Debit dan Pencegahan Bencana Banjir di Kali Batan Purwosari Kediri. Neutron 9 vol. 1 pp. 20-31. 2009.

[14] PSDA Ciliwung – Cisadane. Kegiatan Manajemen DPS dan Hidrologi Balai Pendayagunaan Sumber Daya Air Wilayah Sungai Ciliwung – Cisadane. Laporan. Kalibrasi Bendung Ciliwung – Katulampa. Bogor. 2004.

[15] E.R. Amien. "Analisis Pola Sebaran Curah Hujan di Daerah Aliran Sungai Cisadane," Institut Pertanian Bogor. 2016.

[16] Y.S. Swarinoto and Sugiyono. "Pemanfaatan Suhu Udara dan Kelembapan Udara dalam Persamaan Regresi untuk Simulasi Prediksi Total Hujan Bulanan di Bandar Lampung (The Use of Air Temperature and Relative Humidity Into Regression Equation to Provide Simulation Predictions of Monthly Rainfall Total at Bandar Lampung). Metheorology and Geophysics Vol. 12. No. 3 pp. 271-281, December 2011.

[17] B. Santosa, Suharyanto, D. Legono. "Penerapan Optimasi Parameter pada Metode Exponential Smoothing untuk Perkiraan Debit." Media Komunikasi Teknik Sipil. Tahun 18 No.1, February 2010.

[18] S. Singgih. "Menguasai Statistik Multivariat" PT. Elex Media Komputindo. Jakarta. 2015.

[19] Soewarno. Hidrologi: Aplikasi Metode Statistik untuk Analisis Data. NOVA Publishing. Bandung.

[20] Sudjana. Metoda Statistika. Tarsito Publishing. Bandung. 1995.

[21] Sugiyono. Statistika untuk Penelitian. Alfabeta, Bandung. 2005

[22] S. Supriyadi. "Prediksi Total Hujan Bulanan di Tanjungpandan Menggunakan Persamaan Regresi dengan Prediktor SST Nino 3.4 dan India Ocean Dipole (IOD) (Prediction of Rain Total Monthly in Tanjungpandan Using Regression Equation with Predictors SST Nino 3.4 and India Ocean Dipole (IOD)). Journal of Mathematics, Saint, dan Teknologi, Vol. 16, No. 2, pp. 1 – 8, September 2015.

[23] B. Triatmodjo. Hidrologi Terapan. Yogyakarta. Beta Offset Yogyakarta. 2013

[24] C. Trihendrari. “Step By Step SPSS 13, Analisis Data Statistik”.Andi Publishing. Yogyakarta. 2013.

[25] M. Wibowo. “Pemodelan Statistik Hubungan Debit dan Kandungan Sedimen Sungai Contoh Kasus di DAS Citarum – Nanjung.” Journal of Environmental Technology Vol. 2 No. 3 pp. 255-260, 2001.

[26] J. L. Wikerson, “Regional regression equation to estimate syntetic unit hydrograph parameters for indiana” Purdue University, West Lafayette, Indiana, 2009.

[27] B.X. Xu, “Regional Flood Frequency Analysis for Southwestern Alberta,” Departement of Civil Engineering, The University of Calgary. Canada, 1999.